Master's Exam - Fall 2008

October 2, 2008
1:00 pm – 5:00 pm

Number
~~NAME:~~ ________________________________

A. The number of points for each problem is given.

B. There are 12 problems with varying numbers of parts.

1 – 5	Probability (125 points)
6 – 12	Statistics (135 points)

C. Write your answers on the exam paper itself. If you need more room you may use the extra sheets provided. Answer as many questions as you can on each part. Tables are provided.

PROBABILITY PART.

1. (a) (7 pts.) A group of 32 boys and 28 girls is randomly divided into 30 pairs. Find the probability that each girl will be paired with a boy.

(b) (7 pts.) If 20 children are randomly selected from a group of 32 boys and 28 girls, what is the probability that there will be 5 girls selected?

2. Suppose that 20% of newborn boys and 10% of newborn girls are left-handed. Assume that 50.5% of a population of newborns are boys and 49.5% are girls.

(a) (3 pts.) Find the probability that a randomly selected newborn is left-handed.

(b) (3 pts.) Given that a randomly selected newborn is not left-handed, what is the conditional probability that the selected newborn was a boy?

(c) (6 pts.) Suppose that a random sample of 80 newborns is selected. Find (approximate) probability that there were 2 left-handed newborn girls in the sample.

(d) (6 pts.) Suppose that a random sample of 2000 newborns is selected. Find (approximate) probability that there were 80 left-handed newborn girls in the sample.

3. (a) (3 pts.) Let X_n have the cdf F_n for $n = 1, 2, ...$ Let X have the cdf F. Define: X_n converges in distribution to X as $n \to \infty$.

(b) (10 pts.) Let $X_1, ..., X_n, ...$ be independent random variables such that X_n takes value 0 with probability $p_n = \frac{1+n}{2n}$ and value $(1 + 1/n)^n$ with probability $1 - p_n$. Show that X_n converges in distribution to some X as $n \to \infty$. Find the distribution of X.

(c) (3 pts.) Define: X_n converges in probability to 1 as $n \to \infty$.

4. Consider random variables X and Y, jointly distributed with the density function

$$f(x,y) = \begin{cases} \frac{4}{3\pi} & , 1 < x^2 + y^2 < 4, x > 0, y > 0 \\ 0 & , \text{otherwise.} \end{cases}$$

(a) (7 pts.) Find the marginal density of X. (Hint: graph the area of integration.)

(b) (10 pts.) Find the conditional density of Y given $X = x$, such that $1 < x < 2$.

(c) (10 pts.) Find $Z = \mathrm{E}(Y|X)$.

(d) (5 pts.) Find $V = \mathrm{E}(X + X^2 + 5|X)$.

5. Let X_1, X_2, X_3 be independent random variables with densities $f_1(x) = 2x, f_2(x) = 3x^2, f_3(x) = 1$ for $0 \le x \le 1$ correspondingly.

(a) (5 pts.) Given a random variable U, uniformly distributed on $[0, 1]$, how could such a random variable X_2 be generated from U?

(b) (10 pts.) Let $M = \min(X_1, X_2, X_3)$. Find the cdf and density of M.

(c) (10 pts.) Let $S = X_1 - X_2$. Find the cdf of S. (Hint: graph the area of integration).

(d) (8 pts.) Find the cdf and the density for $Y = \ln(1 + X_1^2)$.

(e) (6 pts.) Find the covariance $\mathrm{Cov}(6X_1 - 10X_2 + 3X_3, -3X_1 + 5X_2 + 6X_3)$.

(f) (6 pts.) Find $E(W)$ for

$$W \quad \sqrt{X_1}X_2^3$$

STATISTICS PART.

6. Let σ be a unknown parameter. Let $(X_1, \cdots, X_n)$ be a random sample from the density

$$f(x) = \frac{1}{\sigma}e^{-(x-1)/\sigma}, x > 1.$$

(a) (10 pts.) Find the maximum likelihood estimator (mle) of σ.

(b) (10 pts.) Find the method of moments estimator of σ.

(c) (5 pts.) Find the asymptotic variance $1/nI(\sigma)$ of the mle.

7. Let $f_0(x) = 3x^2, 0 \le x \le 1$ and $f_1(x) = 5x^4, 0 \le x \le 1$. Suppose that we wish to test $H_0 : X$ has density f_0 vs $H_a : X$ has density f_1.

(a) (8 pts.) Construct the most powerful test of $\alpha \in (0, 1)$ significance level based on X.

(b) (3 pts.) Which theorem did you use in part (a)?

(c) (7 pts.) Find the power of this test.

8. (12 pts.) The following table provides data on body weight gain (g) for two independent samples: a sample of animals given a 1 mg/pellet dose of a certain soft steroid and a sample of control animals:

Steroid	28.0	29.9	39.3	38.0	29.3	34.3
Control	28.8	30.4	36.4	30.2	28.7	24.2

State a nonparametric model, define hypotheses, and carry out a test at level $\alpha = 0.1$ which will enable you to decide whether there is a significant difference in weight gain between steroid and control groups.

9. How does energy intake compare to energy expenditure? To study this issue, a random sample of professional basketball players was taken. Below the results are summarized in a table (MJ/day):

Player	1	2	3	4	5	6	7
Expenditure	14.0	9.1	12.6	13.7	14.7	13.8	10.4
Intake	9.6	11.0	18.4	16.1	16.5	12.1	16.0

(a) (12 pts.) State a parametric model, define hypotheses, and carry out a test at level $\alpha = 0.1$ which will enable you to decide whether there is a significant difference between intake and expenditure. Find the p-value for this test.

(b) (8 pts.) Perform a nonparametric test of the same hypotheses with $\alpha = 0.1$. Find the p-value for this test.

10. Let (x_i, Y_i) be observed for $i = 1, \cdots, n$. Suppose that the $x_i \neq 0$ are constants, $Y_i = \beta\sqrt{x_i} + \varepsilon_i$, where β is a unknown parameter and the ε_i are independent, each with the $N(0, \sigma^2)$ distribution.

(a) (8 pts.) Show that the least squares estimator for β is $\hat{\beta} = (\sum_{i=1}^{n} Y_i\sqrt{x_i})/(\sum_{i=1}^{n} x_i)$. Do **NOT** use matrix or vector space methods.

(b) (5 pts.) Show that $\hat{\beta}$ is an unbiased estimator of β. Do **NOT** use matrix or vector space methods.

(c) (6 pts.) Find $Var(\hat{\beta})$. Do **NOT** use matrix or vector space methods.

(d) (5 pts.) Suppose that each $\sqrt{x_i}$ is replaced by $\sqrt{cx_i}$, where $c \neq 0$ is a constant. Let $\hat{\beta}^*$ be the new least squares estimator for β. Find an expression of $\hat{\beta}^*$ in terms of $\hat{\beta}$.

11. Let p_1 and p_2 be the proportions of a population of 45 thousand MSU students who wanted a decrease in tuition at time t_1 and time t_2, respectively. In order to estimate $\Delta = 3p_1 - 2p_2$ random samples of sizes $n_1 = 50$ and $n_2 = 100$ were taken independently at the two times. Let X_1 and X_2 be the numbers favoring a decrease in tuition for the two samples.

(a) (8 pts.) Define notation and give a formula for a 95% confidence interval on Δ.

(b) (5 pts.) Apply the formula for the case $X_1 = 40, X_2 = 90$.

(c) (7 pts.) Explain the meaning of your interval in such a way that someone who had not ever studied statistics would understand.

(d) (6 pts.) If equal sample sizes, $n = n_1 = n_2$ were to be used how large must n be in order that the estimator $\hat{\Delta}$ have probability at least 0.95 of being within 0.1 of Δ?

12. (10 pts.) A random sample of MSU students is obtained, and each individual is categorized with respect to level of studying and political affiliation. Does the accompanying data suggest that there is an association between level of studying and political affiliation?

Affiliation	Democrat	Republican	Independent
Undergraduate	96	67	37
Graduate	57	26	17

State a model, define hypotheses, and carry out a test at level $\alpha = 0.1$.